The Kayser–Cuny Mathematical Tables in Molecular and Synthetic Biology: Foundations of a Molecular Information Framework

Victoria Kayser-Cuny

MTMVP (France)

Abstract

The genetic code is traditionally represented as a correspondence table linking 64 codons to 20 amino acids and three translation stop signals. While this representation accurately describes protein translation, it does not explicitly capture potential mathematical relationships underlying codon organization. Here, we introduce the **Kayser–Cuny Mathematical Tables**, a unified molecular information framework centred on the **Theoretical Codonic Unit (TCU)**, a deterministic hierarchical representation of the genetic code derived from atomic composition. Starting from the atomic invariants of the four constituent elements (C, H, N and O), a single mathematical transformation generates numerical representations of nucleobases, theoretical doublets and codons, leading to a deterministic reorganization of the canonical codon table. Matrix analysis, thresholded heatmaps, graph representations and spectral decomposition consistently reveal recurrent numerical invariants centred on the values 1, 96–97 and 128, together with reproducible topological organizations. As a proof of concept, the numerical descriptors derived from the TCU were further applied to protein sequence design and optimization, yielding de novo and optimized protein sequences with predicted three-dimensional structures reaching AlphaFold pTM confidence scores above 0.90. Together, these results establish the Kayser–Cuny Mathematical Tables as a quantitative framework for investigating the genetic code and suggest potential applications in computational biology, synthetic biology and rational protein engineering.

Keywords Genetic code; Numerical invariants; Codon classification; Boolean algebra; Information theory; Graph theory; Protein engineering; AlphaFold; Synthetic biology; Theoretical Codonic Unit (TCU).

Introduction

The genetic code is generally represented as a correspondence table linking 64 codons to 20 amino acids and three translation stop signals. This canonical representation accurately describes protein translation and has been extensively investigated from biochemical, evolutionary and informational perspectives. Nevertheless, it does not explicitly address whether the organization of the genetic code itself can be described by a unified mathematical framework spanning multiple hierarchical scales.

In this work, we introduce the **Kayser–Cuny Mathematical Tables**, a molecular information framework that reconstructs the organization of the genetic code from the atomic composition of its constituent elements. The framework is centred on the **Theoretical Codonic Unit (TCU)**, a deterministic numerical representation generated from a single mathematical transformation applied recursively from atoms to nucleobases, theoretical doublets and codons.

This hierarchical construction produces a numerical reorganization of the canonical codon table that can be analysed using matrix visualization, graph theory and spectral methods. Rather than replacing the biological representation of the genetic code, the TCU provides a complementary mathematical description that enables quantitative investigation of its internal organization.

We show that this representation consistently reveals recurrent numerical invariants, reproducible topological organizations and characteristic spectral signatures across multiple independent analyses. Finally, as a proof of concept, the numerical descriptors derived from the TCU are applied to protein sequence design and optimization, illustrating the potential of the framework for computational biology and synthetic biology.

2. Methodological Framework: Theoretical Codonic Unit (TCU)

To establish a quantitative representation of the genetic code, we introduce the **Theoretical Codonic Unit (TCU)**, a deterministic mathematical framework that assigns a numerical descriptor to each hierarchical level of molecular organization, from atoms to complete codons.

The construction begins with the four chemical elements constituting DNA nucleobases—hydrogen (H), carbon (C), nitrogen (N) and oxygen (O). For each element, an initial numerical value vvv is defined as the product of the numbers of protons (ppp), neutrons (nnn) and electrons (eee):

$$v = p \times n \times e$$

This value is subsequently transformed using a single nonlinear function applied uniformly throughout the framework:

$$f(v) = \frac{v^3}{4}$$

The division by four reflects the quaternary organization of the genetic code, characterized by four principal chemical elements (H, C, N and O) and four canonical nucleobases (A, T, C and G). No additional coefficients, empirical corrections or optimization procedures are introduced during the construction.

Phosphorus (P), although essential for the phosphodiester backbone of DNA, is intentionally excluded from the initial numerical transformation. Unlike the CHON elements, whose relative abundances differ among adenine, thymine, cytosine and guanine, phosphorus is present in an identical structural context in every nucleotide. Consequently, it does not contribute to distinguishing the informational identity of individual nucleobases. The present framework therefore focuses on the variable atomic composition of carbon, hydrogen, nitrogen and oxygen, which provides the discriminating chemical basis for the hierarchical numerical representation.

The transformed atomic values are subsequently propagated through successive hierarchical levels to generate numerical representations of nucleobases, theoretical doublets and finally the complete set of sixty-four codons. This deterministic procedure produces a numerical reorganization of the canonical genetic code, referred to as the Theoretical Codonic Unit (TCU).

Throughout this work, the TCU serves as the reference numerical framework from which all matrix analyses, network representations, spectral analyses and protein design constraints are derived.

Table 1. Proposed Theoretical Codonic Unit (TCU):
Hierarchical Organization and Atomic Invariants

Gap between column 1	Subatomic particles raised to the cube	Combinaison	Family	Number of particles cubed and divided by 4	Constant numerical patterns - Gap between column 5
	1	Electron	Particle	**0.25 ou 1 ?**	Does not apply.
+26	**27** (3 quarks = 3 x 3 x 3)	Proton	Particle	**6.75** (27/4)	
+0	**27** (3 quarks = 3 x 3 x 3)	Neutron	Particle	**6.75** (27/4)	
1	**1** (1 electron x 1 proton)	Hydrogen	Atom	**0.25 ou 1 ?**	**1**
+215	**216** (6 protons x 6 neutrons x 6 electrons)	Carbon	Atom	**54** (216/4)	+54
+127	**343** (7 protons x 7 neutrons	Ntrogen	Atom	**85.75** (343/4)	+31.75

	x 7 electrons)				
+169 Decompositon: (1+215+169 = **512, voir colonne à côté)**	**512** (8 protons x 8 neutrons x 8 electrons)	Oxygen	Atom	**128** (512/4)	+42.25 (note : 54 + 42.25 = **96.25**, et 54 + 31.75 + 42.25 = **128**)
+1898 Decompositon: (1508 + 396 + 4 = **1898**)	**2410** [$C_4H_5N_3O$ = (216x4) + (1x5) + (343x3) + (512x1) = 2140]	C	Nucleic base	**602.5** (2410/4)	**+474.50** (97.5 + 377)
+386 Decompositon: (382+4 =**386**)	**2796** [$C_5H_6N_2O_2$ = (216x5)+(1X6)+(343x2) +(512x2) = 2796)	T	Nucleic base	**699** (2796/4)	**+96.5**
+4	**2800** [$C_5H_5N_5$ = (216x5)+(1x5)+(343x5) = 2800)	A	Nucleic base	**700** (2800/4)	**+1**
+512 Decompositon: (122+386+4= **512)**	**3312** [$C_5H_5N_5O$ = (216x5)+(1x5)+(343x5) +(512x1) = 3312)	G	Nucleic base	**828** (3312/4)	**+128**
+1508 Decompositon: (1898-386-4 = **1508**)	**4820** (C+C = 2410+2410)	CC	Theoretical ancient code with doublets	**1205** (4820/5)	**+377**
+386 Decompositon: (382+4 = 386)	**5206** (T+C or C+T = 2796 + 2410 = 5206)	TC-CT	Theoretical ancient code with doublets	**1301.5** (5206/4)	**+96.5**
+4 Decompositon: (1508 + 396 + 4 = 1898)	**5210** (A+C or C+A = 2800+2410 = 5210)	AC-CA	Theoretical ancient code with doublets	**1302.5** (5210/4)	**+1**
+382 Decompositon: (386-4= **382**)	**5592** (T+T = 2796+2796 = 5592)	TT	Theoretical ancient code with doublets	**1398** (5592/4)	+95.5
+4	**5596** (A+T or T+A = 2800+2796= 5596)	AT-TA	Theoretical ancient code with doublets	**1399** (5596/4)	+1
+4	**5600** (A+A=2800 +2800)	AA	Theoretical ancient code	**1400** (5600/4)	+1

			with doublets		
+122 Decompositon: (127-4-1 =**122**)	**5722** (C+G or G+C= 2410+3312=5722)	CG-GC	Theoretical ancient code with doublets	**1430.5** (5722/4)	+30.5 (note : 95.5 + 1 + 1 + 30.5 = **128**)
+386 Decompositon: (386-4= **382**)	**6108** (T+G or G+T = 2796+3312 = 6108)	TG-GT	Theoretical ancient code with doublets	**1527** (6108/4)	**+96.5**
+4	**6112** (A+G or G+A = 2800 + 3312=6112)	AG-GA	Theoretical ancient code with doublets	**1528** (6112/4)	**+1**
+512 Decompositon: (122+38 6+4= **512)**	**6624** (G+G = 3312+3312=6624)	GG	Theoretical ancient code with doublets	**1656** (6624/4)	**+128**
+606 Decompositon: (518+21 5-121= **606**)	**7230** (C+C+C=2410+2410+2 410=7230)	CCC	Codon	**1807** (7230/4)	+151
+386 Decompositon: (386-4= **382**)	**7616** (C+C+T or T+C+C or C+T+C = 2410+2410+2796 = 7616)	CCT-TCC-CTC	Codon	**1904** (7616/4)	**+97**
+4	**7620** (C+C+ A or A+C+C or C+A+C = 2410+2410+2800 = 7620)	CCA-ACC-CAC	Codon	**1905** (7620/4)	**+1**
+382 Decompositon: (386-4= **382**)	**8002** (T+C+T or C+T+T or T+T+C = 2796+2410+2796 = 8002)	TCT-CTT-TTC	Codon	**2000.5** (8002/4)	+95
+4	**8006** (T+C+ A or A+C+T or C+T+A or A+T+C or C+A+T or T+A+C = 2796+2410+2800 = 8006)	TCA-ACT-CTA-ATC-CAT-TAC	Codon	**2001** (8006/4)	+1
+4	**8010** (A+C+A or C+A+A or A+A+C = 2800+2410+2800 = 8010)	ACA-CAA-AAC	Codon	**2002.5** (8010/4)	+1

+122 Decompositon: (127-4-1 =**122**)	**8132** (C+C+G or G+C+C or C+G+C = 2410+2410+3312= 8132)	CCG-GCC-CGC	Codon	**2033** (8132/4)	+31 (note : 95 + 1 + 1 + 31 = **128**)
+256 Decompositon: (382-122 -4=**256**)	**8388** (T+T+T = 2796+2796+2796= 8388)	TTT	Codon	**2097** (8388/4)	+64
+4	**8392** (T+T+A or A+T+T or T+A+T = 2796+2796+2800 = 8392)	TTA-ATT-TAT	Codon	**2098** (8392/4)	+1
+4	**8396** (A+T+A or T+A+A or A+A+T = 2800+2796+2800 = 8396)	ATA-TAA-AAT	Codon	**2099** (8396/4)	+1
+4	**8400** (A+A+A = 2800 +2800+2800 = 8400)	AAA	Codon	**2100** (8400/4)	+1
+518 Decompositon: (386+ 127+4+1 =**518**)	**8518** (T+C+G or G+C+T or C+T+G or G+T+C or C+G+T or T+G+C = 2796+2410+3312= 8518)	TCG-GCT-CTG-GTC-CGT-TGC	Codon	**2129.5** (8518/4)	+28.5 (note : 64 + 1 + 1 +1 +28 = **95.5**)
+4	**8522** (A+C+G or G+C+A or C+A+G or G+A+C or C+G+A or A+G+C = 2800+2410+3312= 8522)	ACG-GCA-CAG-GAC-CGA-AGC	Codon	**2130.5** (8522/4)	**+1**
+382 Decompositon: (386-4= **382**)	**8904** (T+T+G or G+T+T or T+G+T = 2796+2796+3312= 8904)	TTG-GTT-TGT	Codon	**2226** (8904/4)	+96
+4	**8908** (A+T+G or G+T+A or T+A+G or G+A+T or T+G+A or A+G+T = 2800+2796+3312= 8908)	ATG-GTA-TAG-GAT-TGA-AGT	Codon	**2227** (8908/4)	+1
+4	**8912** (A+A+G or G+A+A or A+G+A = 2800+2800+3312= 8912)	AAG-GAA-AGA	Codon	**2228** (8912/4)	+1
+122	**9034**	GCG-CGG-	Codon	**2258.5**	+30.5

Decompositon: (127-4-1 =**122**)	(G+C+G or C+G+G or G+G+C = 3312+2410+3312= 9034)	GGC		(9034/4)	(note : 96 + 1 +1 + 30 = **128.5**)
+386 Decompositon: (386-4= **382**)	**9420** (G+T+G or T+G+G or G+G+T = 3312+2796+3312= 9420)	GTG-TGG-GGT	Codon	**2355** (9420/4)	**+97**
+4	**9424** (G+A+G or A+G+G or G+G+A =3312+3312+2800= 9424)	GAG-AGG-GGA	Codon	**2356** (9424/4)	**+1**
+512 Decompositon: (122+386+4= **512**)	**9936** (G+G+G= 3312+3312+3312= 9936)	GGG	Codon	**2484** (9936/4)	**+128**

It is also possible to include the following codons (uracil instead of thymine):

Tableau 2.

UGA	STOP	8872	2218
UAG	STOP	8872	2218
UAA	STOP	8360	2090
AUG	START	8872	2218

UGA, UAG, and AUG have exactly the same logical value. 2218 - 2090 = **128**.

3. Matrix Classification and Symmetry Analysis of the 64 Codons

The numerical reorganization of the sixty-four codons reveals a highly structured matrix organization that is not directly apparent in the canonical genetic code. When codons are reordered according to the **Kayser–Cuny Mathematical Tables**, regular numerical transitions emerge between successive base columns, together with reproducible diagonal and block symmetries.

Unlike the conventional codon table, which is primarily organized according to nucleotide positions for translation, this numerical classification groups codons according to their deterministic values derived from the Theoretical Codonic Unit (TCU). The resulting arrangement preserves the biological identities of the codons while exposing previously hidden numerical relationships.

Inspection of the reordered matrix reveals that the transitions between adjacent base groups are predominantly governed by recurrent numerical increments centred on the invariants **Δ = 1**, **Δ = 96–97** and **Δ = 128**. These invariants appear repeatedly across the matrix and define its principal organizational axes.

Table 2 presents the complete numerical classification of the 64 codons according to the Kayser–Cuny framework.

TABLE OF THE 64 CODONS – KAYSER-CUNY CLASSIFICATION (2026)

1st BASE		2nd BASE: C			+96-97	2nd BASE: T			+1	2nd BASE: A			+128	2nd BASE: G			= between 225–226
C	a	CCC	Proline (Pro)	1807	96	CTC	Leucine (Leu)	1904		CAC	Histidine (His)	1905		CGC	Arginine (Arg)	2033	
	b	CCT	Proline (Pro)	1904	97	CTT	Leucine (Leu)	2000		CAT	Histidine (His)	2001		CGT	Arginine (Arg)	2129	
	c	CCA	Proline (Pro)	1905	97	CTA	Leucine (Leu)	2001		CAA	Glutamine (Gln)	2002		CGA	Arginine (Arg)	2130	
	d	CCG	Proline (Pro)	2033		CTG	Leucine (Leu)	2129		CAG	Glutamine (Gln)	2130		CGG	Arginine (Arg)	2258	
T	a	TCC	Serine (Ser)	1904	96	TTC	Phenylalanine (Phe)	2000		TAC	Tyrosine (Tyr)	2001		TGC	Cysteine (Cys)	2129	
	b	TCT	Serine (Ser)	2000		TTT	Phenylalanine (Phe)	2097		TAT	Tyrosine (Tyr)	2098		TGT	Cysteine (Cys)	2226	
	c	TCA	Serine (Ser)	2001	97	TTA	Leucine (Leu)	2098		TAA	STOP	2099		TGA	STOP	2227	
	d	TCG	Serine (Ser)	2129		TTG	Leucine (Leu)	2226		TAG	STOP	2227		TGG	Tryptophan (Trp)	2355	
A	a	ACC	Threonine (Thr)	1904	97	ATC	Isoleucine (Ile)	2001		AAC	Asparagine (Asn)	2002		AGC	Serine (Ser)	2130	
	b	ACT	Threonine (Thr)	2001		ATT	Isoleucine (Ile)	2098		AAT	Asparagine (Asn)	2099		AGT	Serine (Ser)	2227	
	c	ACA	Threonine (Thr)	2002	97	ATA	Isoleucine (Ile)	2099		AAA	Lysine (Lys)	2100		AGA	Arginine (Arg)	2228	
	d	ACG	Threonine (Thr)	2130		ATG	Methionine (Met)	2227		AAG	Lysine (Lys)	2228		AGG	Arginine (Arg)	2356	
G	a	GCC	Alanine (Ala)	2033	96	GTC	Valine (Val)	2129		GAC	Aspartic acid (Asp)	2130		GGC	Glycine (Gly)	2258	
	b	GCT	Alanine (Ala)	2129		GTT	Valine (Val)	2226		GAT	Aspartic acid (Asp)	2227		GGT	Glycine (Gly)	2355	
	c	GCA	Alanine (Ala)	2130	97	GTA	Valine (Val)	2227		GAA	Glutamic acid (Glu)	2228		GGA	Glycine (Gly)	2356	
	d	GCG	Alanine (Ala)	2258		GTG	Valine (Val)	2355		GAG	Glutamic acid (Glu)	2356		GGG	Glycine (Gly)	2484	

COLOR LEGEND (AMINO ACIDS)

- **Purple :** Nonpolar aliphatic amino acids
- **Blue :** Nonpolar amino acids (branched-chain or aromatic)
- **Green :** Polar uncharged amino acids
- **Red :** Positively or negatively charged (polar) amino acids
- **Red (STOP) :** Stop codons

ΔI = +225.5
(between 225–226)

Calculation of a codon's logical value

$$v_i = p_i \cdot n_i \cdot e_i$$

where:

- v = logical value of a codon
- p = number of protons
- n = number of neutrons
- e = number of electrons

The resulting logical value is then raised to the third power and divided by four to reflect the quaternary organization of the system: four principal atoms (C, H, N and O), four nucleobases (A, C, G and T/U), and the organization of codons into groups of four.

$$f(v) = \frac{v^3}{4}$$

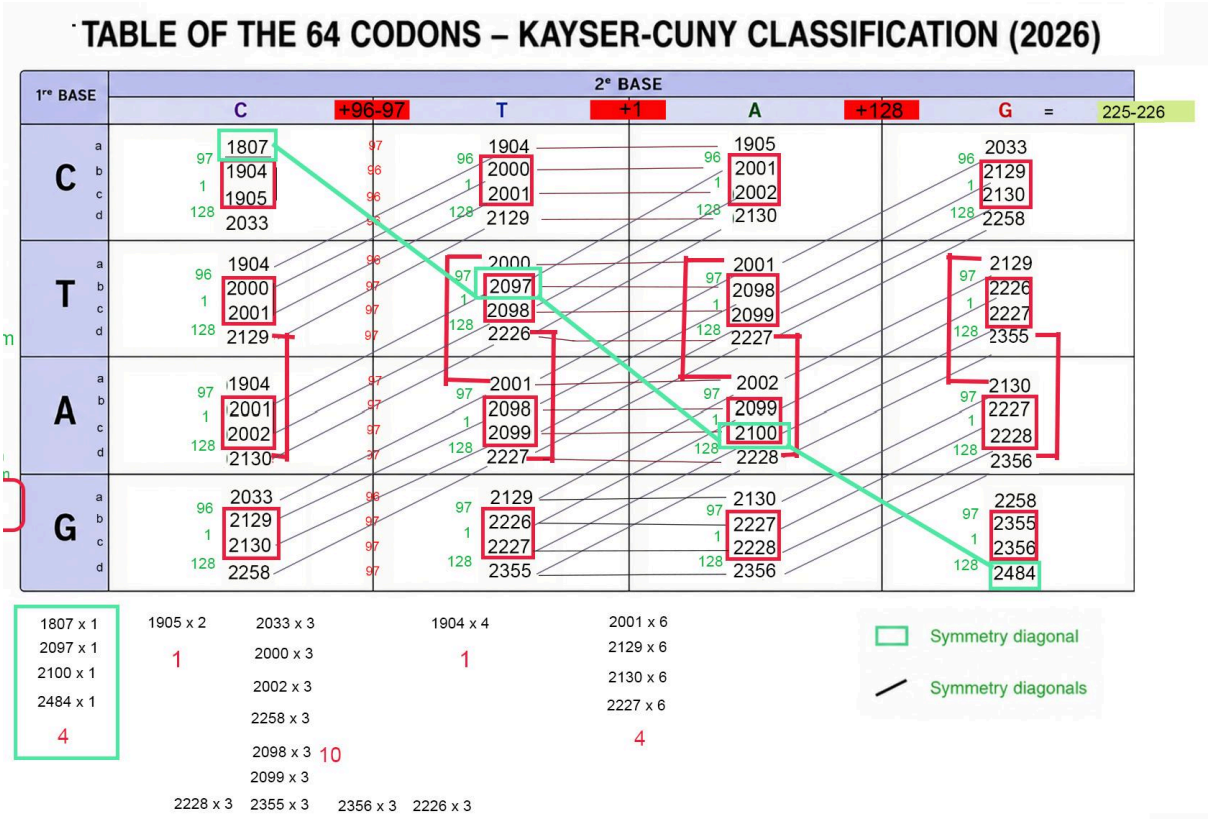
Beyond the numerical ordering itself, the reorganized matrix exhibits several reproducible geometric symmetries. Multiple codon groups form repeated vertical, horizontal and diagonal arrangements, indicating that the numerical organization is not random but follows deterministic structural constraints.

Particularly striking are the long diagonal trajectories crossing the matrix, which connect codons sharing identical or closely related numerical transitions. Several extreme values occupy symmetric positions near opposite corners of the matrix, while intermediate values are distributed into recurrent blocks exhibiting comparable local organizations.

The frequencies of individual numerical values also display a non-uniform distribution. Some values occur uniquely, whereas others appear repeatedly in highly regular multiplicities, generating hierarchical levels within the matrix organization.

These observations collectively suggest that the numerical representation generated by the TCU possesses an intrinsic internal structure that can be further investigated using graph-theoretical and spectral approaches.

Table 3 highlights the principal diagonal symmetries and repeated numerical motifs identified within the reordered codon matrix.



4. Boolean Factorization of Codon Representations

Following the numerical reorganization of the genetic code, each codon was converted into an expanded symbolic representation preserving the four canonical nucleobases (C, T, A and G). The expanded expressions were then systematically simplified using the first distributive law of Boolean algebra.

The objective of this procedure was not to redefine the biological function of codons but to determine whether their numerical organization could be expressed through a reduced logical representation. Each of the sixty-four codons therefore underwent the same deterministic sequence of operations, consisting of symbolic expansion, Boolean factorization and numerical evaluation.

The first distributive law,

$$A(B + C) = AB + AC$$

was applied identically to every codon representation.

The resulting factorized expressions generate a reduced logical form while preserving the numerical descriptors derived from the Theoretical Codonic Unit (TCU). Successive numerical differences between neighbouring codons reveal recurrent transition values

centred on the invariants **Δ = 1**, **Δ = 96–97** and **Δ = 128**, which remain conserved throughout the factorization process.

This procedure therefore provides a deterministic Boolean representation of the reorganized codon matrix and establishes a direct correspondence between symbolic simplification and the numerical invariants identified in the previous section.

Table 3 summarizes the Boolean factorization of the sixty-four codons and the numerical transitions obtained after simplification.

Table 3. Codons interpreted through the prism of the first distributive law of Boolean algebra, applied to binary arithmetic and digital logic circuits.

DNA	Original	Simplified	Results	Number of particles cubed and divided by 4	Gap	Invariants decom-position
CCC	CTAGC CTAGC C	CTAG [C] = CTAGC+CTAGCC [CC]	CTAG (C+CC)	1807		
CCT	CTAGC CTAGC T	CTAG [C] = CTAGC+CTAGCT [CT]	CTAG (C+CT)	1904	+97	
CCA	CTAGC CTAGC A	CTAG [C] = CTAGC+CTAGCA [CA]	CTAG (C+CA)	1905	+1	
CCG	CTAGC CTAGC G	CTAG [C] = CTAGC+CTAGCG [CG]	CTAG (C+CG)	2033	+128	
CTC	CTAGC CTAGT C	CTAG [C] = CTAGC +CTAGTC [TC]	CTAG (C+TC)	1904	-129	(129=1 28+1)
CTT	CTAGC CTAGT T	CTAG [C] = CTAGC +CTAGTT [TT]	CTAG (C+TT)	2000	+96	
CTA	CTAGC CTAGT A	CTAG [C] = CTAGC +CTAGTA [TA]	CTAG (C+TA)	2001	+1	
CTG	CTAGC CTAGT G	CTAG [C] = CTAGC +CTAGTG [TG]	CTAG (C+TG)	2129	+128	

CAC	CTAGC CTAGA C	CTAG [C] = CTAGC +CTAGAC [AC]	CTAG (C+AC)	1905	-224	(224 = 128+96)
CAT	CTAGC CTAGA T	CTAG [C] = CTAGC +CTAGAT [AT]	CTAG (C+AT)	2001	+96	
CAA	CTAGC CTAGA A	CTAG [C] = CTAGC +CTAGAA [AA]	CTAG (C+AA)	2002	+1	
CAG	CTAGC CTAGA G	CTAG [C] = CTAGC +CTAGAG [AG]	CTAG (C+AG)	2130	+128	
CGC	CTAGC CTAGG C	CTAG [C] = CTAGC +CTAGGC [GC]	CTAG (C+GC)	2033	-97	(97= 96+1)
CGT	CTAGC CTAGG T	CTAG [C] = CTAGC +CTAGGT [GT]	CTAG (C+GT)	2129	+96	
CGA	CTAGC CTAGG A	CTAG [C] = CTAGC +CTAGGA [GA]	CTAG (C+GA)	2130	+1	
CGG	CTAGC CTAGG G	CTAG [C] = CTAGC +CTAGGG [GG]	CTAG (C+GG)	2258	128	
TCC	CTAGT CTAGC C	CTAG [T] = CTAGT +CTAGCC [CC]	CTAG (T+CC)	1904	-129	(129=1 28+1)
TCT	CTAGT CTAGC T	CTAG [T] = CTAGT +CTAGCT [CT]	CTAG (T+CT)	2000	+96	
TCA	CTAGT CTAGC A	CTAG [T] = CTAGT +CTAGCA [CA]	CTAG (T+CA)	2001	+1	
TCG	CTAGT CTAGC G	CTAG [T] = CTAGT +CTAGCG [CG]	CTAG (T+CG)	2129	+128	
TTC	CTAGT CTAGT C	CTAG [T] = CTAGT +CTAGTC [TC]	CTAG (T+TC)	2000	-129	(129=1 28+1)

TTT	CTAGT CTAGT T	CTAG [T] = CTAGT +CTAGTT [TT]	CTAG (T+TT)	2097	+97	
TTA	CTAGT CTAGT A	CTAG [T] = CTAGT +CTAGTA [TA]	CTAG (T+TA)	2098	+1	
TTG	CTAGT CTAGT G	CTAG [T] = CTAGT +CTAGTG [TG]	CTAG (T+TG)	2226	+128	
TAC	CTAGT CTAGA C	CTAG [T] = CTAGT +CTAGAC [AC]	CTAG (T+AC)	2001	-225	(128+97=225)
TAT	CTAGT CTAGA T	CTAG [T] = CTAGT +CTAGAT [AT]	CTAG (T+AT)	2098	+97	
TAA	CTAGT CTAGA A	CTAG [T] = CTAGT +CTAGAA [AA]	CTAG (T+AA)	2099	+1	
TAG	CTAGT CTAGA G	CTAG [T] = CTAGT +CTAGAG [AG]	CTAG (T+AG)	2227	+128	
TGC	CTAGT CTAGG C	CTAG [T] = CTAGT +CTAGGC [GC]	CTAG (T+GC)	2129	-98	(98=97+1)
TGT	CTAGT CTAGG T	CTAG [T] = CTAGT +CTAGGT [GT]	CTAG (T+GT)	2226	+97	
TGA	CTAGT CTAGG A	CTAG [T] = CTAGT +CTAGGA [GA]	CTAG (T+GA)	2227	+1	
TGG	CTAGT CTAGG G	CTAG [T] = CTAGT +CTAGGG [GG]	CTAG (T+GG)	2355	+128	
ACC	CTAGA CTAGC C	CTAG [A] = CTAGA +CTAGCC [CC]	CTAG (A+CC)	1904	-451	(451=97+1 +128 +97+ 128)

ACT	CTAGACTAGCT	CTAG [A] = CTAGA +CTAGCT [CT]	CTAG (A+CT)	2001	+97	
ACA	CTAGACTAGCA	CTAG [A] = CTAGA +CTAGCA [CA]	CTAG (A+CA)	2002	+1	
ACG	CTAGACTAGCG	CTAG [A] = CTAGA +CTAGCG [CG]	CTAG (A+CG)	2130	+128	
ATC	CTAGACTAGTC	CTAG [A] = CTAGA +CTAGTC [TC]	CTAG (A+TC)	2001	-129	(129=128+1)
ATT	CTAGACTAGTT	CTAG [A] = CTAGA +CTAGTT [TT]	CTAG (A+TT)	2098	+97	
ATA	CTAGACTAGTA	CTAG [A] = CTAGA +CTAGTA [TA]	CTAG (A+TA)	2099	+1	
ATG	CTAGACTAGTG	CTAG [A] = CTAGA +CTAGTG [TG]	CTAG (A+TG)	2227	+128	
AAC	CTAGACTAGAC	CTAG [A] = CTAGA +CTAGAC [AC]	CTAG (A+AC)	2002	-225	(225=128+97)
AAT	CTAGACTAGAT	CTAG [A] = CTAGA +CTAGAT [AT]	CTAG (A+AT)	2099	+97	
AAA	CTAGACTAGAA	CTAG [A] = CTAGA +CTAGAA [AA]	CTAG (A+AA)	2100	+1	
AAG	CTAGACTAGAG	CTAG [A] = CTAGA +CTAGAG [AG]	CTAG (A+AG)	2228	+128	
AGC	CTAGACTAGGC	CTAG [A] = CTAGA +CTAGGC [GC]	CTAG (A+GC)	2130	-98	(98=97+1)

AGT	CTAGA CTAGG T	CTAG [A] = CTAGA +CTAGGT [GT]	CTAG (A+GT)	2227	+97	
AGA	CTAGA CTAGG A	CTAG [A] = CTAGA +CTAGGA [GA]	CTAG (A+GA)	2228	+1	
AGG	CTAGA CTAGG G	CTAG [A] = CTAGA +CTAGGG [GG]	CTAG (A+GG)	2356	+128	
GCC	CTAGG CTAGC C	CTAG [G] = CTAGG +CTAGCC [CC]	CTAG (G+CC)	2033	-323	(323= 96+1+ 128+9 7+1)
GCT	CTAGG CTAGC T	CTAG [G] = CTAGG +CTAGCT [CT]	CTAG (G+CT)	2129	+96	
GCA	CTAGG CTAGC A	CTAG [G] = CTAGG +CTAGCA [CA]	CTAG (G+CA)	2130	+1	
GCG	CTAGG CTAGC G	CTAG [G] = CTAGG +CTAGCG [CG]	CTAG (G+CG)	2258	+128	
GTC	CTAGG CTAGT C	CTAG [G] = CTAGG +CTAGTC [TC]	CTAG (G+TC)	2129	-129	(129= 128+ 1)
GTT	CTAGG CTAGT T	CTAG [G] = CTAGG +CTAGTT [TT]	CTAG (G+TT)	2226	+97	
GTA	CTAGG CTAGT A	CTAG [G] = CTAGG +CTAGTA [TA]	CTAG (G+TA)	2227	+1	
GTG	CTAGG CTAGT G	CTAG [G] = CTAGG +CTAGTG [TG]	CTAG (G+TG)	2355	+128	
GAC	CTAGG CTAGA C	CTAG [G] = CTAGG +CTAGAC [AC]	CTAG (G+AC)	2130	-225	(128+ 97)

GAT	CTAGG CTAGA T	CTAG [G] = CTAGG +CTAGAT [AT]	CTAG (G+AT)	2227	+97	
GAA	CTAGG CTAGA A	CTAG [G] = CTAGG +CTAGAA [AA]	CTAG (G+AA)	2228	+1	
GAG	CTAGG CTAGA G	CTAG [G] = CTAGG +CTAGAG [AG]	CTAG (G+AG)	2356	+128	
GGC	CTAGG CTAGG C	CTAG [G] = CTAGG +CTAGGC [GC]	CTAG (G+GC)	2258	-98	(98=97 +1)
GGT	CTAGG CTAGG T	CTAG [G] = CTAGG +CTAGGT [GT]	CTAG (G+GT)	2355	+97	
GGA	CTAGG CTAGG A	CTAG [G] = CTAGG +CTAGGA [GA]	CTAG (G+GA)	2356	+1	
GGG	CTAGG CTAGG G	CTAG [G] = CTAGG +CTAGGG [GG]	CTAG (G+GG)	2484	+128	

5. Numerical Codon Classification for Rational Protein Design

The deterministic numerical organization generated by the **Kayser–Cuny Mathematical Tables** was subsequently extended from codon analysis to protein sequence construction. Rather than using the canonical codon table as the primary design framework, codons were reorganized according to their numerical groups derived from the Theoretical Codonic Unit (TCU).

Each codon was assigned to a numerical class defined by the recurrent invariants identified throughout the previous analyses. Codons sharing identical numerical descriptors were grouped together independently of their conventional arrangement within the standard genetic code. This restructuring generated a numerical coding system in which synonymous and non-synonymous codons could be explored according to their shared mathematical properties.

The resulting classification constitutes a deterministic design space for protein engineering. Instead of selecting codons exclusively according to amino acid identity, sequences can be constructed or modified while simultaneously preserving predefined numerical constraints derived from the TCU.

Table 5 summarizes the numerical classification of codons used throughout the protein design procedure.

Table 5. Numerical restructuring of codons for protein encoding according to the Kayser–Cuny Mathematical Tables.

Groupe	Codon	Amino acids
1	CCA	Proline
1	ACC	Threonine
1	CAC	Histidine
1	TCA	Serine
1	ACT	Threonine
1	CTA	Leucine
1	ATC	Isoleucine
1	CAT	Histidine
1	TAC	Tyrosine
1	ACA	Threonine
1	CAA	Glutamine
1	AAC	Asparagine
1	TTA	Leucine
1	ATT	Isoleucine
1	TAT	Tyrosine
1	ATA	Isoleucine
1	TAA	STOP
1	AAT	Asparagine
1	AAA	Lysine
1	ACG	Threonine
1	GCA	Alanine
1	CAG	Glutamine
1	GAC	Aspartic acid
1	CGA	Arginine

1	AGC	Serine
1	ATG	Methionine START
1	GTA	Valine
1	TAG	STOP
1	GAT	Aspartic acid
1	TAG	STOP
1	GAT	Aspartic acid
1	TGA	STOP
1	AGT	Sérine
1	AAG	Lysine
1	GAA	Glutamic acid
1	AGA	Arginine
1	GAG	Glutamic acid
1	AGG	Arginine
1	GGA	Glycine

Groupe	**Codon**	**Amino acids**
28.5	TCG	Serine
28.5	GCT	Alanine
28.5	CTG	Leucine
28.5	GTC	Valine
28.5	CGT	Arginine
28.5	TGC	Cysteine

Groupe	**Codon**	**Amino acids**
30.5	GCG	Alanine
30.5	CGG	Arginine
30.5	GGC	Glycine

Groupe	Codon	Amino acids
31	CCG	Proline
31	GCC	Alanine
31	CGC	Arginine

Groupe	Codon	Amino acids
64	TTT	Phenylalanine

Groupe	Codon	Amino acids
95	TCT	Serine
95	CTT	Leucine
95	TTC	Phenylalanine

Groupe	Codon	Amino acids
96	TTG	Leucine
96	GTT	Valine
96	TGT	Cysteine

Groupe	Codon	Amino acids
97	CCT	Proline
97	TCC	Serine
97	CTC	Leucine
97	GTG	Valine
97	TGG	Tryptophane
97	GGT	Glycine

Groupe	Codon	Amino acids

128	GGG	Glycine
128	UAG	Stop
128	UAA	Start

6. Biological Validation Using the Human Erythrocyte Protein System

To evaluate the biological relevance of the **Kayser–Cuny Mathematical Tables**, the numerical framework was applied to a well-characterized set of human erythrocyte proteins, including **hemoglobin**, **α- and β-spectrins**, and the **Band 3 anion transport protein**. These proteins were selected because their primary sequences, tertiary structures and functional domains have been extensively characterized experimentally, providing a suitable reference system for assessing the proposed numerical representation.

Protein sequences were analysed according to the numerical codon classification derived from the **Theoretical Codonic Unit (TCU)**. The numerical descriptors identified throughout the codon analysis were propagated to the amino acid sequences, allowing recurrent invariant groups to be mapped onto the corresponding protein structures.

Within this dataset, the invariant groups centred on **Δ = 96** and **Δ = 97** were repeatedly observed in regions associated with major structural transitions and mechanically important domains. Similar numerical organizations were identified in hemoglobin and in proteins contributing to the erythrocyte membrane skeleton, suggesting that the hierarchical numerical framework remains coherent across proteins with distinct biological functions.

As a proof of concept, the numerical descriptors generated by the TCU were incorporated into the protein design workflow developed in this study. De novo and optimized protein sequences constructed under these numerical constraints were subsequently evaluated using **AlphaFold**, yielding predicted three-dimensional structures with **predicted Template Modeling (pTM) confidence scores exceeding 0.90** for several representative sequences.

These preliminary results indicate that the Kayser–Cuny Mathematical Tables provide a consistent numerical framework capable of guiding protein sequence organization while producing structurally coherent models. Additional experimental validation will be required to determine the biological significance and general applicability of these numerical descriptors across larger protein datasets.

7. Discussion

The present work introduces the **Kayser–Cuny Mathematical Tables** as a deterministic numerical framework for representing and reorganizing the genetic code. Unlike conventional representations, which primarily describe the correspondence between codons and amino acids, this framework establishes a hierarchical numerical organization extending from atomic composition to protein sequence design.

A central result of this study is that the same mathematical transformation generates coherent numerical representations across multiple levels of biological organization. Starting from the atomic composition of the four nucleobases, the framework produces a deterministic codon classification that reveals recurrent numerical invariants, reproducible matrix symmetries and characteristic topological organizations. These patterns were consistently recovered using independent mathematical approaches, including matrix analysis, graph theory and spectral decomposition.

The Boolean factorization presented in this work further demonstrates that the reorganized codon matrix can be expressed through simplified logical relationships while preserving the numerical invariants identified by the Theoretical Codonic Unit. Rather than replacing the canonical genetic code, this representation provides a complementary mathematical language for describing its internal organization.

The extension of this framework to protein sequence design represents an additional proof of concept. By restructuring codons according to their numerical classes, the Kayser–Cuny Mathematical Tables define a deterministic search space that can be explored for de novo protein generation and sequence optimization. The structurally coherent AlphaFold predictions obtained for representative sequences suggest that this numerical organization is compatible with biologically plausible protein folds, although broader experimental validation will be required.

The present work has several limitations. The numerical framework has been evaluated on representative protein datasets and computational structural predictions. Future studies should assess its robustness across larger protein families, experimentally validate newly designed proteins and compare its performance with existing computational protein design strategies.

Beyond its immediate applications, the framework introduced here provides a quantitative basis for investigating relationships between molecular composition, codon organization and protein architecture. Because the methodology is entirely deterministic and reproducible, it may also prove useful for future developments in computational biology, synthetic biology and machine-learning approaches applied to molecular design.

8. Conclusion

The **Kayser–Cuny Mathematical Tables** introduce a unified numerical framework for analysing the genetic code from atomic composition to protein design. Using a single deterministic mathematical transformation, the framework generates hierarchical numerical representations of nucleobases, codons and protein-encoding sequences.

The resulting reorganization of the canonical codon table reveals reproducible numerical invariants, matrix symmetries, logical factorization patterns and characteristic network organizations that are consistently observed across multiple independent mathematical analyses.

As a proof of concept, this numerical framework was successfully applied to the construction and optimization of protein sequences evaluated with AlphaFold, yielding highly confident

predicted structures for representative examples. Although additional experimental validation remains necessary, these results demonstrate that the proposed numerical descriptors can serve as practical constraints for computational protein engineering.

Taken together, these findings establish the **Kayser–Cuny Mathematical Tables** as a deterministic molecular information framework that complements the canonical genetic code and provides new mathematical tools for quantitative investigations in molecular biology, synthetic biology and rational protein design.

References

Foundations of DNA and the Genetic Code

1. WATSON, J., CRICK, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738 (1953). https://doi.org/10.1038/171737a0
2. CHARGAFF E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia. 1950 Jun 15;6(6):201-9. doi: 10.1007/BF02173653. PMID: 15421335.
3. NIRENBERG MW, MATTHAEI JH. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A. 1961 Oct 15;47(10):1588-602. doi: 10.1073/pnas.47.10.1588. PMID: 14479932; PMCID: PMC223178.
4. Crick FH. The origin of the genetic code. J Mol Biol. 1968 Dec;38(3):367-79. doi: 10.1016/0022-2836(68)90392-6. PMID: 4887876.
5. Crick F. Central dogma of molecular biology. Nature. 1970 Aug 8;227(5258):561-3. doi: 10.1038/227561a0. PMID: 4913914.
6. Khorana HG, Büchi H, Ghosh H, Gupta N, Jacob TM, Kössel H, Morgan R, Narang SA, Ohtsuka E, Wells RD. Polynucleotide synthesis and the genetic code. Cold Spring Harb Symp Quant Biol. 1966;31:39-49. doi: 10.1101/sqb.1966.031.01.010. PMID: 5237635.

Information Theory and Boolean Algebra

7. Shannon, C.E. (1948), A Mathematical Theory of Communication. Bell System Technical Journal, 27: 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
8. Boole, G. (1854) An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities. Walton and Maberly, London, 184. https://doi.org/10.5962/bhl.title.29413
9. Gillie, A. C. Binary Arithmetic and Boolean Algebra. McGraw-Hill Book Company, New York, 1965.

Mathematical Representations of the Genetic Code

10. He MX, Petoukhov SV, Ricci PE. Genetic code, hamming distance and stochastic matrices. Bull Math Biol. 2004 Sep;66(5):1405-21. doi: 10.1016/j.bulm.2004.01.002. PMID: 15294430.
11. The Genetic Code, Hadamard Matrices and Algebraic Biology. Matthew He () and Sergey Petoukhov (). Journal of Biological Systems 2010 18:spec01, 159-175 10.1142/S0218339010003688
12. Rosandić M, Paar V. The Supersymmetry Genetic Code Table and Quadruplet Symmetries of DNA Molecules Are Unchangeable and Synchronized with Codon-Free Energy Mapping during Evolution. Genes (Basel). 2023 Dec 12;14(12):2200. doi: 10.3390/genes14122200. PMID: 38137022; PMCID: PMC10743133.

Evolution and Optimization of the Genetic Code

18. Haig D, Hurst LD. A quantitative measure of error minimization in the genetic code. J Mol Evol. 1991 Nov;33(5):412-7. doi: 10.1007/BF02103132. Erratum in: J Mol Evol 1999 Nov;49(5):708. PMID: 1960738.
19. Freeland SJ, Hurst LD. The genetic code is one in a million. J Mol Evol. 1998 Sep;47(3):238-48. doi: 10.1007/pl00006381. PMID: 9732450.

20. Knight RD, Freeland SJ, Landweber LF. Selection, history and chemistry: the three faces of the genetic code. Trends Biochem Sci. 1999 Jun;24(6):241-7. doi: 10.1016/s0968-0004(99)01392-4. PMID: 10366854.
21. Freeland SJ, Knight RD, Landweber LF, Hurst LD. Early fixation of an optimal genetic code. Mol Biol Evol. 2000 Apr;17(4):511-8. doi: 10.1093/oxfordjournals.molbev.a026331. PMID: 10742043.
22. Massey SE. A neutral origin for error minimization in the genetic code. J Mol Evol. 2008 Nov;67(5):510-6. doi: 10.1007/s00239-008-9167-4. Epub 2008 Oct 15. PMID: 18855039.
23. Koonin EV, Novozhilov AS. Origin and evolution of the genetic code: the universal enigma. IUBMB Life. 2009 Feb;61(2):99-111. doi: 10.1002/iub.146. PMID: 19117371; PMCID: PMC3293468.
24. Koonin EV. Frozen Accident Pushing 50: Stereochemistry, Expansion, and Chance in the Evolution of the Genetic Code. Life (Basel). 2017 May 23;7(2):22. doi: 10.3390/life7020022. PMID: 28545255; PMCID: PMC5492144.

Graph Theory, Networks and Spectral Analysis

25. Chung, F.R.K. (1997) Spectral Graph Theory. American Mathematical Society, Providence.
26. Newman, M. (2010) Networks: An Introduction. Oxford University Press, Oxford. http://dx.doi.org/10.1093/acprof:oso/9780199206650.001.0001

Artificial Intelligence, Protein Structure Prediction and Protein Design

29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583-589. doi: 10.1038/s41586-021-03819-2. Epub 2021 Jul 15. PMID: 34265844; PMCID: PMC8371605.
30. Protein complex prediction with AlphaFold-Multimer. Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, Demis Hassabis. bioRxiv 2021.10.04.463034; doi: https://doi.org/10.1101/2021.10.04.463034
31. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues AV, van Dijk AA, Ebrecht AC, Opperman DJ, Sagmeister T, Buhlheller C, Pavkov-Keller T, Rathinaswamy MK, Dalwadi U, Yip CK, Burke JE, Garcia KC, Grishin NV, Adams PD, Read RJ, Baker D. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021 Aug 20;373(6557):871-876. doi: 10.1126/science.abj8754. Epub 2021 Jul 15. PMID: 34282049; PMCID: PMC7612213.
32. Anishchenko I, Pellock SJ, Chidyausiku TM, Ramelot TA, Ovchinnikov S, Hao J, Bafna K, Norn C, Kang A, Bera AK, DiMaio F, Carter L, Chow CM, Montelione GT, Baker D. De novo protein design by deep network hallucination. Nature. 2021 Dec;600(7889):547-552. doi: 10.1038/s41586-021-04184-w. Epub 2021 Dec 1. PMID: 34853475; PMCID: PMC9293396.

33. Watson, J.L., Juergens, D., Bennett, N.R. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023). https://doi.org/10.1038/s41586-023-06415-8
34. Adleman LM. Molecular computation of solutions to combinatorial problems. Science. 1994 Nov 11;266(5187):1021-4. doi: 10.1126/science.7973651. PMID: 7973651.
35. Endy D. Foundations for engineering biology. Nature. 2005 Nov 24;438(7067):449-53. doi: 10.1038/nature04342. PMID: 16306983.

Complete Corpus

Kayser-Cuny, V. (2025). Meta-Genesis. Towards a Biology without Matter, based on Pure Logic. Multi-Scale Numerical Invariants and Fractal Properties of the Genetic Code (Abstract and compilation). Zenodo. https://zenodo.org/records/21002033

Kayser-Cuny, V. (2025). (Part 1) Multi-Scale Numerical Invariants and Fractal Properties of the Genetic Code: A Combinatorial and Atomic Analysis. Zenodo. https://zenodo.org/records/21002648

Kayser-Cuny, V. (2025). (Part 2) Multiscale Numerical Invariants and Fractal Properties of the Genetic Code: Internal Constraints and Multiscale Packet Distributions Revealing a Universal Grammar. Zenodo. https://doi.org/10.5281/zenodo.17272500

Kayser-Cuny, V. (2025). (Part IV-part 3) Multi-Scale Numerical Invariants and Fractal Properties of the Genetic Code: A Unified Theory of Biological Information, from Stars to Codons. Zenodo. https://doi.org/10.5281/zenodo.17370443

Kayser-Cuny, V. (2025). (Part VI-part 3) Multi-Scale Numerical Invariants and Fractal Properties of the Genetic Code: Data Availability [Data set]. Zenodo. https://doi.org/10.5281/zenodo.17306204

Kayser-Cuny, V. (2025). Data Availability Part 2 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.17368936

Kayser-Cuny, V. (2025). (Part III) The Mirror-Twin Paradox: A New Approach to DNA Understanding the Implications of an Inverted Genome and Its Applications in Molecular Genetics, Neuroscience, and Medicine. Zenodo. https://doi.org/10.5281/zenodo.15390489

Kayser-Cuny, V. (2025). Meta-Genesis. Towards a Biology Without Matter. From Boolean Algebra to the Expansion of Life: Binary Arithmetic and Multi-Dimensional Projections of the Genetic Code. Zenodo. https://doi.org/10.5281/zenodo.17494922

A Deterministic Method for the Generation, Simulation, and Assembly of De Novo Proteins Based on Numerical Invariants Intrinsic to the Genetic Code:

Part 1. Kayser-Cuny, V. (2026). The Kayser–Cuny Mathematical Tables in Molecular and Synthetic Biology: A Molecular Information Framework (2026th ed.). MTMVP. https://zenodo.org/records/21131611 ISBN: 9782489162035

Part 2. Kayser-Cuny, V. (2026). PROOF OF CONCEPT Multi-scale Numerical Invariants and Fractal Properties of the Genetic Code: A Combinatorial and Atomic Analysis using the Erythrocyte (Red Blood Cell) as an Ideal Mathematical Model for AI-Based Proteomic Analysis. Zenodo. https://doi.org/10.5281/zenodo.21003215

All rights reserved.

Copyright Victoria Kayser-Cuny.

Editor: MTMVP (France)

ISBN 978-2-489162-03-5

Dépôt légal BnF: Juillet 2026

No part of this publication may be reproduced, distributed, or transmitted, in any form or by any means, without the prior written permission of Victoria Kayser-Cuny or the publisher MTMVP.

Print-on-demand for environmental reasons: Lulu.com.